

ARC920010085US1
10/042,367In the Claims:

1. (Currently Amended) A method of determining parameter combinations for automated web crawler access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, said method comprising:

maintaining at least one log file containing ~~at least one set of parameters~~ user queries resulting from previous real user HTML interactions with said World Wide Web site; analyzing said log file to determine parameter combinations ~~for~~ and to generate synthetic queries for input to said web crawler, said web crawler using said input for automated access to said World Wide Web content.

2. (Currently Amended) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 1, ~~wherein said parameters are entries in HTML forms, said user queries comprising entries~~, said analyzing step further comprising

ranking entries ~~in each set of entries~~ according to their frequency of occurrence; for ~~each~~ a set of entries resulting from unlimited text entries, excluding entries ranked below a predetermined number; and

wherein said ~~parameter combination~~synthetic queries are determined by producing combinations of entries from each set of entries.

3. (Currently Amended) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 2, ~~wherein said~~ synthetic queries -

ARC920010085US1
10/042,367

parameter combinations are determined by producing all combinations of entries from each set of entries.

4. (Original) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 2, wherein entries resulting from limited text entries and unlimited text entries have stop words removed and remaining words stemmed.

5. (Original) A method of determining parameter combinations for automated access to World Wide Web content that is accessible based on parameters resulting from real user interactions with a World Wide Web site, as per claim 1, wherein said log file is maintained by a proxy server that logs communications between a client and a Web server resulting from real user accesses to said World Wide Web content.

6. (Canceled)

7. (Currently Amended) A method of increasing web crawler penetration of Web databases accessible via HTML forms, said method comprising:

reviewing previous real user form input data;

identifying possible HTML form input data for said Web crawler from said previous real user form input data by synthesis of entries for any of: predefined sets, limited text entries or unlimited text entries; and

providing said identified form input data to said Web crawler during an instantiation of automated access to said Web databases by said Web crawler.

ARC920010085051
10/042,367

8. (Previously Presented) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 7, wherein said previous form input data are maintained in a log file.

9. (Original) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 8, wherein said log file is maintained by a proxy server.

10. (Original) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 7, wherein said synthesis comprises:

ranking any entries for predetermined sets;
ranking any entries for limited text entries;
ranking any entries for unlimited text entries;
excluding entries for unlimited text entries ranked below a predetermined number;

and

pairing entries from each set of ranked entries.

11. (Original) A method of increasing web crawler penetration of Web databases accessible via HTML forms, as per claim 10, wherein said synthesis further comprises:

removing stop words and stemming remaining words for entries resulting from limited text entries and unlimited text entries.

12. (Canceled)

ARC9200100850S1
10/042,367

13. (Canceled)

14. (Canceled)

15. (Canceled)

16. (Canceled)

17. (Currently Amended) A method of emulating real user access to World Wide Web content dynamically accessible via an HTML form, said method comprising:

maintaining a log containing real user entries into each input item of said HTML form;

ranking entries for each input item according to their frequency of occurrence; for each unlimited text entry input item, excluding entries ranked below a predetermined number;

determining combinations of entries from each set of entries; and automatically accessing said content using said combinations of entries as HTML input for a webcrawler.

18. (Original) A method of emulating real user access to World Wide Web content dynamically accessible via an HTML form, as per claim 17, wherein entries resulting from limited text entries and unlimited text entries have stop words removed and remaining words stemmed.

ARC920010085US1
10/042,367

19. (Original) A method of emulating real user access to World Wide Web content dynamically accessible via an HTML form, as per claim 17, wherein said log file is maintained by a proxy server that logs communications between a client and a Web server resulting from real user accesses to said World Wide Web content.

20. (Canceled)

21. (Canceled)

22. (Canceled)

23. (Canceled)

24. (Canceled)

25. (Canceled)